·论著·全科医学教育研究·

程序性评价在全科临床诊疗思维课程评价中的应用

翟佳燚[1，2，3]，陆媛[1，3，4*]，石建军[1，3，4]，于德华[1，3，4]

1.200090 上海市，同济大学附属杨浦医院全科医学科
2.201899 上海市嘉定区嘉定镇街道社区卫生服务中心
3.200090 上海市，同济大学医学院全科医学系
4.200090 上海市全科医学与社区卫生发展研究中心
*通信作者：陆媛，主任医师；Email：lussy@126.com

【摘要】 **背景** 全科临床诊疗思维是全科医生为患者提供高质量诊疗服务的关键，但因其属于无法被直接观察的抽象能力，所以并不适合用传统的方法来对该能力进行评价。**目的** 本文旨在通过优化课程教学评价方法来提高全科医生临床诊疗思维培训效率、丰富医学教育评价。**方法** 于2022年9~12月，以同济大学医学院38名全科方向硕士研究生为对象，根据生源类别分为在职组16人与住培组22人。课程包含核心、案例、结果3个模块，采取混合式教学方法开展实践。定量评价的内容为技能评分表、PBL评分表与案例评分表，主观评价的内容来自多源反馈。采用SPSS 22.0软件进行统计分析。**结果** 两组学员核心和案例模块首次课程评分无统计学意义（P>0.05），但两个模块末次课程得分有统计学意义（P<0.05）。将全班以及两组学员在核心和案例模块首、末次课程的定量评价得分进行比较，均有统计学意义（P<0.001）。主观评价性质分为肯定性评价与改进性评价，发现随着课程的深入，核心模块与案例模块的肯定性评价数量与百分比呈持续上升的趋势，改进性评价数量与百分比则呈持续下降的趋势。**结论** 程序性评价能够丰富全科思维评价体系、推动学员循序建构全科思维，同时也发现了“翻转课堂”是程序性评价的呈现形式。

【关键词】 全科医学；教学；教育，研究生；程序性评价；临床诊疗思维；翻转课堂
【中图分类号】 R-05 【文献标识码】 A DOI：10.12114/j.issn.1007-9572.2023.0743

Effects of Programmatic Assessment in Clinical Reasoning Courses in General Practice

ZHAI Jiayi[1，2，3]，*LU Yuan*[1，3，4*]，*SHI Jianjun*[1，3，4]，*YU Dehua*[1，3，4]
1.Department of General Practice，Yangpu Hospital of Tongji University，Shanghai 200090，China
2.Community Health Service Center of Jiading Town in Jiading District，Shanghai 201899，China
3.Academic Department of General Practice，Tongji University School of Medicine，Shanghai 200090，China
4.Shanghai General Practice and Community Health Development Research Center，Shanghai 200090，China
**Corresponding author：LU Yuan，Chief physician；Email：lussy@126.com*

【Abstract】 **Background** The clinical reasoning of general practitioners is the key to providing quality medical services to patients，but it is not suitable for traditional methods to evaluate this ability because it is an abstract ability that cannot be directly observed. **Objective** This paper aims to improve the efficiency of general practitioners' clinical reasoning training and enrich the evaluation of medical education by optimizing the evaluation method of curriculum teaching. **Methods** From September to December 2022，38 master's degree students of the Medical School of Tongji University were divided into the in-service group （16） and residential training group （22）. The course consists of three modules，namely core，case and result，and adopts blended teaching method to carry out teaching. The content of quantitative evaluation is different kinds of rating scale，while the

基金项目：中华医学会医学教育分会和中国高等教育学会医学教育专业委员会2020年医学教育研究立项课题（2020B-N08139）——全科硕士研究生临床诊疗思维课程建设与实践；上海市医药卫生发展基金会课题（SE1201933）
引用本文：翟佳燚，陆媛，石建军，等. 程序性评价在全科临床诊疗思维课程评价中的应用［J］. 中国全科医学，2024. DOI：10.12114/j.issn.1007-9572.2023.0743.［Epub ahead of print］［www.chinagp.net］
ZHAI J Y，LU Y，SHI J J，et al. Effects of programmatic assessment in clinical reasoning courses in general practice ［J］. Chinese General Practice，2024.［Epub ahead of print］.



content of subjective evaluation comes from multi-source feedback. SPSS 22.0 was used for statistical analysis. **Results** The final course scores for the core and case modules were statistically significant for both groups（P<0.05），but the first course scores were not（P>0.05）. The quantitative evaluation scores of the whole class and the two groups of students in the first and last courses of the above two modules were compared，and there was statistical significance （P<0.001）. Subjective evaluation can be divided into positive evaluation and improvement evaluation. It is found that with the deepening of the course，the number and percentage of positive evaluation of core modules and case modules continue to rise，while the number and percentage of improvement evaluation continue to decline. **Conclusions** Programmatic assessment can enrich the evaluation system of general clinical reasoning，promote students' construction of general clinical reasoning in order，and also found that "flipped classroom" is a form of procedural evaluation.

【Key words】 General practice；Teaching；Education，graduate；Programmatic assessment；Clinical reasoning；Flipped classroom

全科临床诊疗思维（简称“全科思维”）是一项复合能力，将全科医生问诊、查体、辅助检查运用与解读等专业能力“串连”[1-2]，是为患者提供高质量全科临床诊疗的关键。目前全科思维的培训形式越发多样化[3-5]，如以问题为导向的学习（Problem Based Learning，PBL）和以案例为导向的学习（Case Based Learning，CBL），而全科思维评价作为衡量培训效果的重要手段，也存在着多种评价形式。但是，单一的评价形式难以全面、客观地反映出学员的整体全科思维能力[5-7]。根据米勒金字塔理论[8]，全科思维培养分为“knows”“knows how”“shows how”“dose”4 个阶段，前 3 个阶段的评价被称为“标准化评价”，其通常采用理论考、技能操作考试等成熟方法进行评价；“dose”阶段则需要运用“非标准化评价”[8-9]，由于综合专业知识技能、人文和协作等复合能力被称为“独立于领域的技能”，既往方法难以做到对该能力全面性评价，需要在课程全程采用定性与定量相结合的“非标准化评价”。程序性评价（Programmatic Assessment，PA）兼顾上述两种评价要求，最早由 VAN 等[7]在 2005 年提出，其主要遵循“意义建构”“纵向评价”“持续反馈”“学员主体和自我负责”及“评价为了学习且本身也是学习”五大原则，并强调“评价为了学习且本身也是学习”[9]。PA 结合“形成性评价”促进学员反思和“终结性评价”概括结果的优势[10]，采用“标准化评价”和“非标准化评价”对全科思维进行混合评价，其中任何一个评价方式只作为一个“像素点”是无法全面反映学员整体全科思维的，而当“像素点”达到一定数量以后形成清晰“图像”时才能作出合格与否的决策[11-15]。除了优化教师评价决策，PA 更重要的作用是使学员主动建构专业知识，提高培训效率[11-14]。

《全科临床诊疗思维》是同济大学全科医生岗位胜任能力提升精品课程，也是同济大学全科医学方向两类研究生的共同必修课程，经过多年实践与探索，已取得一定成效[1-2]。本研究通过在同济大学医学院全科医学系的全科医学方向硕士研究生必修课程《全科临床诊疗思维》中引进 PA，旨在探索通过改善评价方法、提高培训效率，为培养高质量全科医生提供思路。

1 对象与方法

1.1 教学对象

将 2022 年参加《全科临床诊疗思维》课程的 38 名学员作为研究对象，其中包括在职全科医生、有全科转岗意愿的专科医生及全科住院医师规范化培训(简称“住培”）医生。根据学习与工作经历分为两组，在职组为研一学术型硕士研究生（简称“在职学员”）共 16 人、住培组为研二“四证合一”专业型硕士研究生（简称“住培学员”）共 22 人，学员基本情况见表 1。

表 1 学员基本情况

Table 1 Students' baseline situation

基本特征	在职组（人）	住培组（人）
性别		
男	8	6
女	8	16
职称		
初级	9	22
中级及以上	7	0
工作年限		
5 年以下	9	22
5 到 10 年	4	0
10 到 20 年	1	0
20 年以上	2	0
全科培训经历		
住培经历	4	22
转岗经历	12	0
全科临床诊疗思维培训经历		
有	8	5
无	8	17
翻转课堂经历		
有	1	6
无	15	16

1.2 课程内容

《全科临床诊疗思维》课程总计32学时，除总论外，分为核心、案例及实践三大模块，采用了理论课、讨论课和实践课等多种形式开展，具体分配课时数见表2。课程的核心模块包含病史采集、体格检查、辅助检查、医患沟通和SOAP病历书写等内容；案例模块包含了涉及健康管理、慢病管理、多病共存、心身疾病和未分化疾病5个方面的全科医学典型案例；结果模块则通过案例汇报的形式考核学员全科思维，并检验学员全科思维完整度。

表2 课程内容
Table 2 Course content

课程模块	课程内容	课时数	授课形式
总论	总论	2	理论课
核心模块	问诊能力、体格检查、辅助检查、医患沟通、病历书写	10	理论课、实践课
案例模块	健康管理、慢病管理、多病共存、心身疾病、未分化疾病	10	讨论课
结果模块	案例汇报	10	案例汇报

1.3 PA设计

1.3.1 PA设计理念

PA可分为3个模块、10条原则[8-9，12，16]，具体内容见表3。标准化评价模块包含4条原则，分别是"能力特定，并非通用""客观性有别于可靠性""评价中提问的作用大于回答""有效性'内置'"；非标准化评价模块同样包含4条原则，分别是"偏见是专业评判的固有特征""评价的有效性取决于工具的使用者""主观内容的重要性""有效使用反馈"[12，16]。

1.3.1.1 标准化评价模块原则内容阐述及应用

（1）"能力特定，并非通用"原则：指学员在任何一种能力测试中的结果不一定代表其在其他测试中的能力，比如理论成绩好的学员不一定操作优秀，这称之为临床能力的"内容特异性"[12]。PA结合了形成性评价与终结性评价的特点，因此框架设计在纵向评价基础上，终结性评价设置在学员课程结束后，形成性评价设置在课程运行中。此外评价内容需要全面考虑，评价问诊、查体、辅助检查、病历书写等"全科思维"核心能力的同时，也要考察医患沟通能力等内容。

（2）"客观性有别于可靠性"原则：其表明即使像客观化结构化临床考试（Objective Structured Clinical Examination，OSCE）这类具有结构化、客观性的特点评价，也同样面临可靠性的问题，试题的全面性才是关键，在该情况下主观评价同样也具有可靠性[12]，因此在课程中应设置直接观察评估和现场点评等主观评价方式。

（3）"评价中刺激的作用大于回应"原则：其指出决定评价内容的是"刺激"而非"回应"。"刺激"指评价形式，如提问、考核、测试等；"回应"指学员答题的过程和结果[12]。这项原则提示了评价设计中应该重视"刺激"的质量和形式，如采用真实案例和场景。本课程中案例模块均来自基层诊室全科工作案例[1-2]，能够增强"刺激"的真实体验。在形式方面，核心与案

表3 程序性评价设计
Table 3 Programmatic assessment design

模块	原则	设计步骤
标准化评价	（1）能力特定，并非通用	（1）纵向评价确立； （2）评价形式设计：形成性与终结性评价； （3）评价内容设计：核心能力、人文沟通
	（2）客观性有别于可靠性	（4）评价方法设计：评分表等定量评价； （5）增加评价方法：主观评价
	（3）评价中提问的作用大于回答	（6）授课形式设计：采用PBL、CBL； （7）增加评价内容：课堂表现评价
	（4）有效性"内置"	（8）课前评价设计：课前课后调查； （9）师资质量控制：组建课程团队，开展师资培训； （10）课程质量控制：邀请专家听课
非标准化评价	（5）偏见是专业评判的固有特征	（11）评价质量控制：评价师资至少设置2人； （12）增加授课形式：增强师生课堂互动
	（6）评价有效性取决于工具使用者	（13）师资质量控制：师资能力持续提升
	（7）主观内容的重要性	（14）增加评价内容：自主学习能力
	（8）有效使用反馈	（15）改进评价方法：主观评价采用多源反馈； （16）课程质量控制：对反馈问题持续追踪
总体评价	（9）没有完美的评价方法	（17）课程质量控制：综合定量和主观评价； （18）课程质量控制：分析所有"像素点"
	（10）评价可驱动学习	（19）课程效果评估：师生访谈和评价结果； （20）课程预期达成：建构全科临床诊疗思维



例模块采用 PBL 与 CBL 等“翻转课堂”形式开展，过程中再进行课堂表现评价，进一步丰富“刺激”形式。

（4）“有效性‘内置’”原则：该原则强调了评价与课程质量改进措施，包括评价前做好设计、评价中做好指引、评价后最好进行分析总结[12，16]。

1.3.1.2 非标准化评价模块原则、总体评价内容阐述及应用

（1）“偏见是专业评判的固有特征”原则：强调了无论是何种评价，偏见是始终存在的，但应该减少偏见对综合专业评判的影响[12，16]。评价中为了减少偏见，遵循三角测量原则，一方面每次定量评价采用 2 名师资进行评价，形成小型决策委员会；另一方面采用了“翻转课堂”，增加师生互动评价，改变刻板印象，减少偏见产生。

（2）“评价的有效性取决于工具的使用者”原则：强调了师资能力应该有持续提升的过程[12]，为此课程组建了设计团队与师资团队。设计团队由 6 名副高级以上全科医学和医学教育领域的专家组成，在课前了解学员的基本情况和需求，课后收集学员反馈意见，在总体框架不变的情况下对课程与评价进行适当改进，同时邀请同行专家对课程与评价方案进行评议，保证质量。师资团队由 6 名具有中级以上职称、5 年以上教学经验的全科师资组成，在课前进行统一培训、统一试讲。

（3）“主观内容重要性”及“有效使用反馈”原则：其提到了当反馈成为评估的核心时，如果师生互动是有效反馈的中介，那么定量信息就有明显局限性，而主观的评价内容则具有优势[12]。回顾评价内容，核心能力、沟通与课堂表现均可采用定量评价与主观评价相结合的方法。根据 PA 中“意义建构”原则，希望学员能够自己建构全科思维，因此增加学习能力评价。在框架设计阶段，核心模块和案例模块最后均设置了反馈环节，并且以提高学员下一次表现为目的，因此对于反馈的方式再一次进行改进，集合了学员自评、学员他评、教师评价、SP 评价等多源反馈（Multi-source Feedback，MSF）。

（4）总体评价：该部分包含了“没有完美的评价方法”和“评价可驱动学习”原则，其对于评价设计的启示是课程评价中所有内容尽可能定量与主观评价相结合，先“打分”后“点评”。

1.3.2 基于“像素点”理论的评价内容设计

基于 PA 的“像素点”理论[11-14]，结合课程核心、案例和结果“三大模块”，将全科思维核心能力、沟通能力、课堂表现和自主学习能力等评价内容细化，评价每个课程的内容，形成定量评价与主观评价结合的“像素点”。定量评价为各模块百分制量化评分表，定性评价为各模块 MSF 中主观评价，分为肯定性评价与改进性评价[17]，肯定性评价指针对学员表现亮点而作出的评价，改进性评价指针对学员表现不足需要改进而作出的评价。最终“像素点”集合反映学员完整全科思维的“图像”。

表 4 评价内容设计
Table 4 Assessment content design

课程模块	课程内容	评价内容	
		定量评价	主观评价
核心模块	问诊能力、体格检查、辅助检查、医患沟通、病历书写	自制评分表	逻辑性、针对性、完整性、符合全科特色、沟通能力
案例模块	健康管理、慢病管理、多病共存、心身疾病、未分化疾病	自制评分表	诊疗技巧、处置合理、全科环境契合度、资源调度、费用控制
结果模块	案例汇报	案例评分表	选题情况、内容价值、汇报风采、课件质量、回答问题

1.4 教学实践过程

于 2022 年 9~12 月，对 38 名全科方向的硕士研究生开展教学实践，课程前向全体学员充分告知研究内容并取得同意。首次课程为总论课程，用于和所有学员以及师资团队成员介绍课程概况，同时也说明了课程的评价形式与师生配合情况，为后续 PA 开展预设基础，在总论课后，对全体学员进行基本情况调查。

核心模块分为知识巩固和实践两部分。在教师引导复习知识以后，由学员扮演医生在标准化病人（Standardized Patients，SP）身上进行实践，同时设置了另一位学员扮演辅助角色，当学员遇到困难时辅助角色可以帮助其共同解决问题。在实践过程中，由教师、其他学员与 SP 对实践学员打分，同时在课程结束后进行 MSF。

案例模块采用线上授课形式，每个案例分 2 次进行授课，每次 1 个课时：首次授课内容为 PBL 讨论，第二次授课内容为学员汇报。课前教师将学员分为 6 组，每组 6 至 7 人，每次 PBL 案例提前 1 至 2 天下发。首次授课时各组先通过“头脑风暴”提出问题，随后由授课老师总结问题并分组认领，学员于课后讨论认领问题并准备汇报材料，每组由组长组织，授课老师通过线上方式参与讨论，每组推选 1 名学员在第二次授课中负责汇报讨论内容，汇报后并由授课师资点评和总结知识点。该模块由教师和组长对所有学员的讨论和课堂表现均进行打分，最后通过 MSF 对学员进行现场反馈。

结果模块为线上案例汇报，学员自行准备一个具有全科特色的病例进行汇报，汇报材料提前 1 周上交，由教师进行 A 至 C 级等级评价，汇报分 3 次进行，根据等级评价结果决定汇报顺序。A 级优先汇报，B 级与 C 级预留时间并参考 A 级学员汇报，结合现场汇报及点评情况进行修改通过后再行汇报。现场汇报由 2 位教师

进行打分，汇报结束后进行 MSF。

1.5 统计学方法

采用SPSS 22.0软件对定量评价结果进行统计分析，数据均符合正态分布，计量资料采用以（$\bar{x} \pm s$）表示，计数资料以构例比表示，率的比较采用卡方检验，组间和前后比较采用 t 检验。以 P<0.05 为差异有统计学意义。

2 结果

2.1 定量评价结果

在本次课程的定量得分中，将两组学员核心、案例和结果 3 个模块的定量评价得分结果进行比较了解两组学员成绩的差异性，结果发现核心和案例模块首次课程评分无统计学意义（P>0.05），核心和案例模块末次课程得分有统计学意义（P<0.05）。将全班以及两组学员在核心和案例模块首、末次课程的定量评价得分进行比较，均有统计学意义（P<0.001）。

表 5 定量评价结果

Table 5 Quantitative evaluation result between the two groups

模块	总体得分	在职组	住培组	t	P
核心模块					
首次课	65.18 ± 11.61	63.71 ± 10.31	66.20 ± 12.60	−0.609	0.547
末次课	84.79 ± 3.60	86.93 ± 3.97	83.30 ± 2.45	3.297	0.002
t	9.798	8.829	6.16		
P	<0.001	<0.001	<0.001		
案例模块					
首次课	65.59 ± 11.79	62.57 ± 12.74	64.30 ± 11.35	−0.416	0.681
末次课	84.82 ± 5.91	88.14 ± 4.87	82.50 ± 5.54	3.069	0.004
t	9.787	8.257	6.38		
P	<0.001	<0.001	<0.001		
结果模块					
案例汇报	85.82 ± 4.27	87.86 ± 4.37	84.40 ± 3.68	2.499	0.018

2.2 主观评价结果

根据转录结果统计数量与百分比情况发现，随着课程的深入，核心模块与案例模块的肯定性评价数量与百分比呈持续上升的趋势，改进性评价数量与百分比则呈持续下降的趋势。

表 6 主观评价结果

Table 6 Qualitative evaluation result

课程模块	肯定性评价		改进性评价	
	数量	百分比（%）	数量	百分比（%）
核心模块				
问诊能力	8	22.86	27	77.14
体格检查	15	31.25	33	68.75
辅助检查	17	41.46	24	58.54
医患沟通	24	61.54	15	38.46
病历书写	29	69.05	13	30.95
案例模块				
健康管理	18	36.73	31	63.27
慢病管理	22	44.90	27	55.10
多病共存	25	54.35	21	45.65
心身疾病	24	58.54	17	41.46
未分化疾病	23	58.97	16	41.03
结果模块				
案例汇报	43	53.75	37	46.25

3 讨论

3.1 PA 能丰富全科思维评价体系

根据本课程的学员定量与主观评价，发现 PA 反映学员全科思维较全面。以往对于全科思维的评价往往采用了病例报告、理论考试、OSCE、临床操作技能评估（Direct Observation of Procedural Skills，DOPS）、迷你临床演练（Mini-Clinical Evaluation Exercise，Mini-CEX）等评价，教学考核和组织实施形式单一，教学评价手段匮乏，导致了评价结果与学员的真实思维水平差距较大[1-5]。本课程有别于既往传统评价，在定量评价的基础上融入主观评价，对教学效果开展混合评价，丰富了全科思维评价体系。与人们对主观评价存在信度不足的"刻板"印象不同，本次评价采用学员自评、教师与学员双向评价、学员互评、SP 对学员的评价组合而成的多源反馈（MSF），符合"三角测量"原则，从而得出了与 VAN 一致的结论，主观评价在评价高级抽象能力时同样可靠[7]。因此，以往理论无法衡量技能水平、"插胃管"等临床操作也难以反映查体技能水平等题目抽样的问题，PA 采用混合评价能够完整地反映学员全科思维，定量评价起到直观评价的作用、主观评价起到了补充与侧面评价的作用。

3.2 "翻转课堂"是 PA 的呈现形式

有别于传统教学，"翻转课堂"是以学员为主体，教师及其他同学作为学习指导和促进者来帮助学习者知识内化，强调了自主学习、交流反思和协作的运用[18]。本次教学对象为全科方向硕士，具备一定的专业基础，因此教师主要起到引导作用而非传授基础理论，帮助学员建构自身全科思维。核心模块中问诊、查体等均是全科医生接诊能力的"基本功"，为后续全科思维将各项能力"串连"打下基础。核心模块主要是技能为主的内容，因此融入实践内容能够让学员得到快速提升，但实践的重点并不是加强技能的熟练度，而是实践运用技能的"思维"，因此采用案例的形式进行实践，同时为了避免学员因能力不足无法达到预期效果，同时设置其他学员担任辅助角色协助处理问题，既是评价反馈又是训练。案例模块均来自具有代表性的全科真实案例，该模

块是提升“串连”能力的主要路径，从PA理论角度来看，也是“串连”能力建构的主要过程。案例模块采用PBL形式，但评价形式采用形成性纵向评价，即在每个课程内容结束后进行评价，观察学员在整个模块培训期间的全科思维发展情况。在PA设计中，课程中形成性评价需要对反馈环节进行调整，以促进学员在下一个课程内容中的表现为目标。最后，案例汇报反映学员“串连”能力建构的结果。该模块同样需要融入反馈，只有通过持续动态评价反馈，才能让学员逐渐形成主动学习意识，并将现有基础知识进一步建构形成全科思维。课程在充分遵循“意义建构”“纵向评价”“持续反馈”“学员主体和自我负责”“评价为了学习且本身也是学习”等PA五大原则的同时，最终以“翻转课堂”形式呈现“案例、实践、讨论、反馈”等内容。

3.3 PA推动学员循序建构全科思维

在PA中，任何一种评价结果都无法决定学员的最终结果，因此对于最终合格与否的决策而言，单一的评价结果是“低利害”决策[6-9]，所有定量与主观评价的集合的结果则是“高利害”决策。生源异质性问题导致的学员定量评分差异在既往研究中已被证实[1-2]，本次研究中，两组核心与案例模块首次课程得分未体现差异，而末次课程中有差异，这可能与PA决策的“利害性”有关。以往核心与案例模块无论课程前后在职学员表现均优于住培学员，主要由实践经验导致。但从授课形式、课程适应度等角度出发，两组学员面对的均是未知的情境，因此导致了本次课程前两组学员表现无差异，随后在课程进行过程中，在职学员经验优势逐渐占优，最终所有模块的表现均优于住培学员。

在核心与案例模块首、末次课定量评价比较中以及主观评价中历次课程肯定性评价与改进性评价数量，能够发现学员表现逐步提升，一定程度能够体现PA能够推动学员对已有知识逐渐内化，同时在全科理念的灌输以后，建构出了全科思维。以往小组讨论、师徒式带教、理论授课、门诊带教等是常用的全科思维培训模式[1-5]，不同的教学模式适用于学员不同能力的培养[5]，但没有一种教学模式是能够全面提升学员所有能力的。全科思维是立足于理论知识与临床技能，又需要将理论与技能“串连”并灵活运用的能力，因此全科思维抽象又涉及多方面能力，需要融合多种模式来教学实践。在本次PA实践中，课前基线调查、课中形成性评价与课后解决性评价相结合以保证课程质量；PBL与CBL的设计，甚至CBL实践中创新加入辅助角色、帮助学员，则是遵循PA“评价为了学习且本身也是学习”原则。

反馈虽是制约课程质量的关键，目前多存在“事前”“事中”反馈缺失、学员知识达成度缺失的问题[19]。本课程中的反馈与以往形成性评价的反馈不同，除了促进学员反思，还能通过利用PA继续评价学员在后续课程中同类问题的改进情况，从而推动学员学习。课程反馈通过MSF实现课程“事中”反馈，同时也反馈给学员知识达成度，促进课程的不断改进、质量的不断提高。因此，从主观评价的数量分析能够发现随着课程的深入，核心模块与案例模块的肯定性评价数量与百分比呈持续上升的趋势，改进性评价数量与百分比则呈持续下降的趋势。在PA理念下的“翻转课堂”，学员通过课程中师生互动、学员实践与合作，得到反馈，最后学员在本身专业知识和技能基础上建构全科思维。

4 小结

本课程遵循了PA与米勒金字塔的理论原则，对同济大学全科方向硕士必修课《全科临床诊疗思维》评价进行优化，在课程中引进PA并开展教学实践，发现了PA能够丰富全科思维评价体系、推动学员循序建构全科思维，同时也发现了“翻转课堂”是PA的呈现形式。本研究也存在不足之处：主观评价内容限于篇幅未能完全展示，后续将进一步完善，并再次对PA的应用效果进行实证。

作者贡献：翟佳燚负责文章构思与设计、研究的实施与可行性分析、撰写论文；陆媛负责文章的英文的修订、文章的质量控制及审校；石建军负责研究的实施；于德华负责论文的修订、监督管理。

本文无利益冲突。

翟佳燚：https://orcid.org/0009-0007-3891-435X

陆媛：https://orcid.org/0000-0002-4360-8476

石建军：https://orcid.org/0009-0003-3117-779X

于德华：https://orcid.org/0000-0001-7652-938X

参考文献

[1] 翟佳燚，陆媛，钱曙蕾，等.同济大学全科硕士研究生临床诊疗思维课程设计[J].中国全科医学.DOI：10.12114/j.issn.1007-9572.2022.0715.

[2] 翟佳燚，陆媛，葛许华，等.基于胜任力的全科临床诊疗思维课程设计与实施[J].中华医学教育杂志，2022，42（12）：1079-1082. DOI：10.3760/cma.j.cn115259-20220502-00566.

[3] 易江，魏涛，于志渊，等.全科住院医师规范化培训中临床思维训练体系建立初探[J].中国毕业后医学教育，2020，4（4）：311-314，324. DOI：10.3969/j.issn.2096-4293.2020.04.006.

[4] 陆媛，于德华，张斌，等.PBL教学模式在全科医师规范化培训中的实践应用[J].中国全科医学，2014，17（16）：1880-1883. DOI：10.3969/j.issn.1007-9572.2014.16.019.

[5] 战京燕，娄景秋，王少坤，等.我国全科住院医师规范化培训教学模式及应用效果研究[J].中国全科医学，2021，24（19）：2401-2407. DOI：10.12114/j.issn.1007-9572.2021.00.201.

[6] VAN DER VLEUTEN C，VERHOEVEN B. In-training assessment developments in postgraduate education in Europe[J]. ANZ J

Surg，2013，83（6）：454-459. DOI：10.1111/ans.12190.

[7] VAN DER VLEUTEN C P M，SCHUWIRTH L W T. Assessing professional competence：from methods to programmes [J]. Med Educ，2005，39（3）：309-317. DOI：10.1111/j.1365-2929.2005.02094.x.

[8] MILLER G E. The assessment of clinical skills/competence/performance [J]. Acad Med，1990，65（9 Suppl）：S63-S67. DOI：10.1097/00001888-199009000-00045.

[9] VAN DER VLEUTEN C P M，DANNEFER E F. Towards a systems approach to assessment [J]. Med Teach，2012，34（3）：185-186. DOI：10.3109/0142159X.2012.652240.

[10] 闻德亮. 医学教育中的形成性评价方法[J]. 高校医学教学研究:电子版，2011，1（1）：36-38.

[11] TORRE D M，SCHUWIRTH L W T，VAN DER VLEUTEN C P M. Theoretical considerations on programmatic assessment [J]. Med Teach，2020，42（2）：213-220. DOI：10.1080/0142159X.2019.1672863.

[12] VAN DER VLEUTEN C P，SCHUWIRTH L W，SCHEELE F，et al. The assessment of professional competence：building blocks for theory development [J]. Best Pract Res Clin Obstet Gynaecol，2010，24（6）：703-719. DOI：10.1016/j.bpobgyn.2010.04.001.

[13] VAN DER VLEUTEN C P，SCHUWIRTH L W，DRIESSEN E W，et al. A model for programmatic assessment fit for purpose [J]. Med Teach，2012，34（3）：205-214. DOI：10.3109/0142159X.2012.652239.

[14] VAN DER VLEUTEN C P M，SCHUWIRTH L W T，DRIESSEN E W，et al. Twelve Tips for programmatic assessment [J]. Med Teach，2015，37（7）：641-646. DOI：10.3109/0142159X.2014.973388.

[15] SCHUWIRTH L W T，VAN DER VLEUTEN C P M. A history of assessment in medical education [J]. Adv Health Sci Educ Theory Pract，2020，25（5）：1045-1056. DOI：10.1007/s10459-020-10003-0.

[16] JOHN A. DENT，RONALD M. HARDEN，DAN HUNT. 医学教师必读——实用教学指导（第6版）[M]. 王维民等，译. 北京：北京大学医学出版社，2020：282-289.

[17] MONICA VAN DE RIDDER J M，STOKKING K M，MCGAGHIE W C，et al. What is feedback in clinical education? [J]. Med Educ，2008，42（2）：189-197. DOI：10.1111/j.1365-2923.2007.02973.x. [LinkOut]

[18] 张金磊，王颖，张宝辉. 翻转课堂教学模式研究 [J]. 远程教育杂志，2012，30（4）：46-51. DOI：10.15881/j.cnki.cn33-1304/g4.2012.04.008.

[19] 丁才成，陈炳和. 高职院校课程质量诊断与改进策略 [J]. 职业技术教育，2017，38（8）：55-58. DOI：10.3969/j.issn.1008-3219.2017.08.019.

（收稿日期：2023-11-14；修回日期：2024-05-14）
（本文编辑：程圣）